Identification and evolution of a novel instructor gene of sex determination in the haplodiploid wasp Nasonia

Zou, Yuan

DOI:
10.33612/diss.134366133

Chapter 5

General discussion

In the preceding chapters, the identification, characterization, and phylogenetic history of the *wasp overruler masculinization* (*wom*) instructor gene for sexual development of the haplodiploid wasp *Nasonia vitripennis* was presented. The existence of this instructor gene was predicted based on previous studies of the sex-determination system of this wasp, but its identity remained unknown. Predictions about its expression pattern and primary function were tested and validated by my research: the instructor gene is expressed from the paternally provided genome in diploid embryos only, and timely activates zygotic *transformer* (*tra*) expression. Activation of *tra* then leads to the production of a functional TRA protein in the early zygote and subsequent female development.

Identification of *wom* in *N. vitripennis*

A differentially expressed gene (DEG) analysis of three sets of 2-5 hours post-oviposition (hpo) embryonic transcriptomes, including haploid (male) only, mostly diploid (female), and haploid gynandromorphic, led to the identification of a candidate gene that is indeed expressed in early diploid embryos but not in haploid embryos. Its zygotic transcription starts 2-3 hpo (blastoderm stage), which is earlier than the onset of zygotic expression of *tra*, peaks at 4-5 hpo, and abruptly declines at 6-7 hpo. Allelic expression analyses revealed that it is transcribed only from the paternal allele in diploid embryos. Therefore, it was concluded that this candidate gene is indeed *wom*. Knockdown of *wom* expression in early diploid embryos by parental RNAi resulted in a significant reduction of both *wom* and *tra* expression levels in diploid embryos and those diploid embryos developed as males. Reversely, *tra* pRNAi had no effect on the early zygotic expression of *wom* in diploid embryos. Together, these findings demonstrate that *wom* is the instructor gene in the *N. vitripennis* sex-determination cascade, and is essential for initializing zygotic expression of *tra* leading to female development. The identification of *wom* confirms and refines the Maternal Effect Genomic Imprinting Sex Determination (MEGISD) model for *Nasonia*.

MEGISD model in *Nasonia*

The original description of the MEGISD model was that a maternal effect gene (*msd*) imprints a zygotic female determining sex-determination gene (*zsd*) (Beukeboom 1995; Beukeboom et al., 2007b); *msd* is active during oogenesis but not spermatogenesis. Haploid unfertilized eggs only contain a maternally imprinted *zsd* copy and develop into males, whereas diploid fertilized eggs in addition contain a paternal non-imprinted *zsd* copy and develop into females. The discovery of *wom* identifies it as the *zsd* in the original model.

I showed that *wom* is temporally expressed in early diploid fertilized eggs, but not in haploid unfertilized eggs, that transcription of *wom* is only from the paternal allele and that it is not maternally provided as mRNA. These results are consistent with the regulation of *zsd* in the

MEGISD model as *zsd* is considered to be a zygotic sex-determining gene and inactivated by maternal imprinting. Zygotic *wom* expression starts at a very early embryonic stage (2-3 hpo), suggesting that the maternal imprinting of *wom* occurs during oogenesis, which is also consistent with the MEGISD model that implies a maternally active imprinting gene, *msd*, that acts during oogenesis. Zygotic *wom* expression is essential for female development, as parental RNAi of *wom* results in a shift from diploid female to male development. Taken together, the MEGISD model for *Nasonia* sex determination can now be viewed as follows: *wom*, the instructor gene for female development, is maternally silenced on the maternal genome in the haploid egg. As a consequence, there is no activation of *tra* in the resulting haploid zygote, that, in the absence of a functional TRA protein, will develop into a male. Upon fertilization, a paternal genome is introduced that carries a non-silenced copy of *wom*. As a result, timely expression of *tra* is initiated. Together with TRA from the maternally provided *tra* mRNA, this leads to sufficient functional TRA to induce female development. Note that both the provision of *tra* mRNA and the silencing of *wom* are maternal effects, whereas the erasing of the imprint of *wom* is a paternal effect. The nature of *msd* and the imprinting mechanism remain to be uncovered. In addition to confirming the MEGISD model, the discovery of *wom* provides crucial knowledge that helps to explain puzzling additional observations concerning *Nasonia* sex determination.

Uniparental female offspring in *Nasonia*

In *N. vitripennis*, two cases of uniparental female offspring have been reported (Beukeboom and Kamping, 2005; Kamping et al., 2007), that seem inconsistent with the haplodiploid sex-determination system and the MEGISD model, as no paternal genome appears required for female development. In a mutant strain, collected from Canada, unmated females produced haploid offspring with female phenotypic characteristics, such as dark pigmentation of antennae and legs or even with an ovipositor (Beukeboom and Kamping, 2005; Beukeboom et al., 2007a). Genetic investigation (Kamping et al., 2007) revealed that a putative *gynandromorph* (*gyn*) locus, located on chromosome 4, acts as a maternal factor in combination with a cytoplasmic component, to cause this phenotype. Unfortunately, the genomic region where *gyn* maps has low recombination and high gene density, which has until now impeded its identification. The results presented in this thesis indicate that *gyn* may well be the putative *msd* locus in the MEGISD model. This would imply that *gyn* is responsible for silencing *wom*, probably by histone modification. Indeed, *wom* was found to not be (completely) maternally silenced in HiCD12 as *wom* is expressed in haploid HiCD12 early embryos in a pattern comparable to that of wild-type diploid embryos (**Figure 5.1**).

Sequence comparison between HiCD12 and other *N. vitripennis* strains did neither indicate potential functional differences in the *wom* gene itself nor in its 2.5 kb upstream sequences

(**Figure 5.2**). Therefore, it is unlikely that gynandromorphism in HiCD12 is caused by a mutation of *wom* itself. Instead, *gyn* is hypothesized to be involved in effectuating the maternal silencing of *wom*, and its loss-of-function mutation in HiCD12 leads to the (partial) failure of maternal *wom* imprinting. Combined investigation of *wom* and the *gyn* genomic region may eventually clarify the identity of *gyn*. In addition, a cytoplasmic effect is involved in producing gynandromorphs (Kamping et al., 2007), suggesting that an interactive effect of a nuclear gene and a cytoplasmic factor is responsible for silencing *wom*.

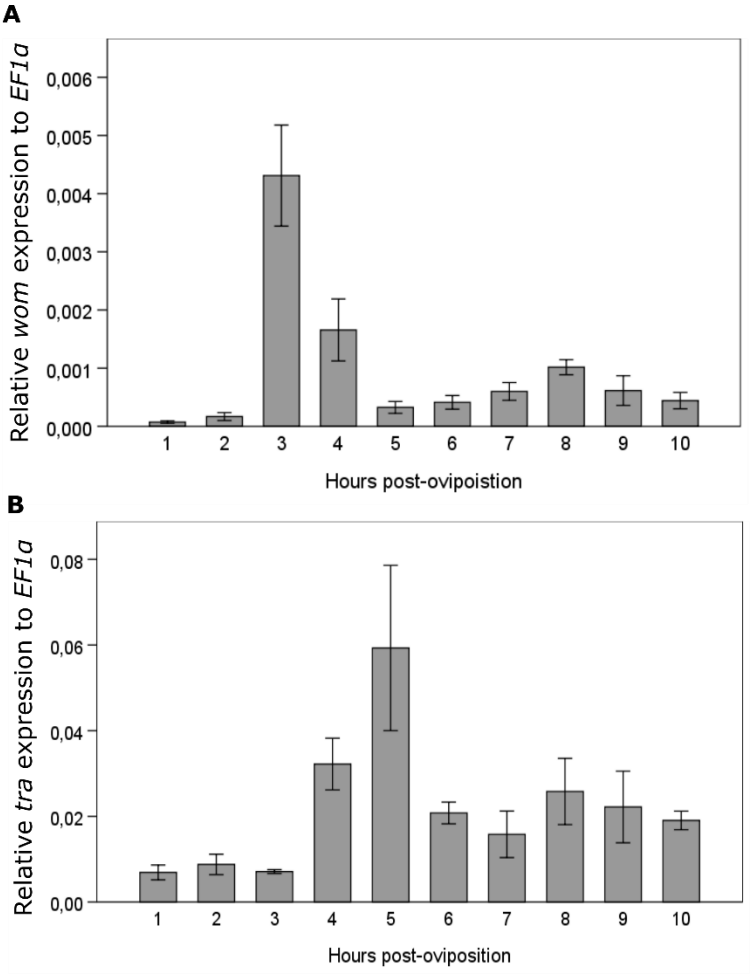
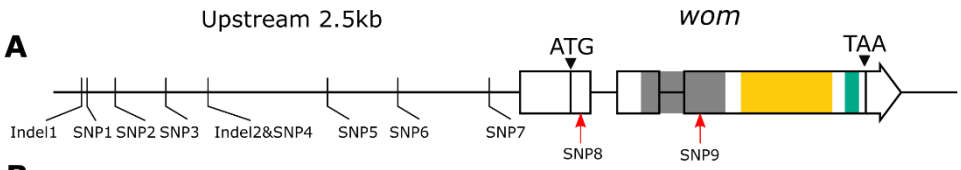


Figure 5.1: Relative temporal expression profiles of *wom* and *tra* in haploid HiCD12 early embryos. Relative expression levels of *wom* (A) and *tra* (B) in early haploid embryos from unmated HiCD12 were normalized to the reference gene *EF1a*. The necessary culturing at 31°C caused the earlier shift of *wom* and *tra* expression peaks. Error bars depict SEM for five biological replicates.



Strains	Indel1	SNP1	SNP2	SNP3	Indel2	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9
HiCD12	--	C	C/A	C/T	CTA/C-A	C/A	A/G	A	GG	C	C
Other strains	CT/--	C/T	C/A	C/T	CTA/C-A	C/A	A/G	A/T	GG/AG	C/T	C/T

Figure 5.2: Comparison of *wom* and its 2.5 kb upstream sequences between HiCD12 and three other *N. vitripennis* strains. (A) Schematic overview of the location of nucleotide variations in *wom* and its upstream sequence. Red arrows indicate two SNPs located in the CDS region. (B) Nucleotide variants between the HiCD12 (Accession: SAMN14532448) and three other strains (Accession: SAMN14532449; PRJNA575073; PRJNA200338). SNP8 is nonsynonymous and SNP9 synonymous. '/' indicates the difference between the upstream regions of two *wom* copies.

A second case of uniparental female offspring has been observed in a polyploid strain (diploid males, triploid females), that has been maintained in the laboratory since it was first described by Whiting (Beukeboom and Kamping, 2005). The fecundity of triploid females is very low due to egg aneuploidy. Unmated triploid females produce some haploid and diploid eggs which normally develop into males, but diploid females can also occasionally be found. This phenomenon appears to be independent of *gyn* (Kamping and Beukeboomn, unpublished). With the current knowledge, I can put forward the following hypothesis about the cause of these uniparental females: the expression level of the gene responsible for *wom* silencing in combination with the cytoplasmic factor (e.g. mitochondria or small RNAs) is not sufficient to fully imprint two *wom* loci in unfertilized diploid eggs. This defective *wom* imprinting may be due to a disruption of gene dosage combined with a mutation in the *wom* gene.

Parent-of-origin effects in insect sex determination

In Hymenoptera, the characteristic of haplodiploid reproduction is that the paternal genome is required for inducing female development. In chapter 3, the zygotic transcription of *wom* in diploid fertilized eggs was shown to initiate from the paternally derived allele only, in line with the hypothesis that *wom* is maternally imprinted, and revealing a parent-of-origin-effect on sex determination in *Nasonia*. This is the first molecular identification of a sex determination imprinting process.

Parent-of-origin-effects on sex determination, such as paternal genome elimination (PGE), have been reported in some other insects, but very little is known about their genetic and molecular basis. Under PGE the entire paternal genome is eliminated or silenced during embryonic development (Normark, 2003). In some Diptera, such as Sciaridae and Cecydomyiidae species, PGE involves the elimination of only the paternal X-chromosomes (Crouse, 1960; Stuart and Hatchett, 1991; Sánchez, 2014). Zygotes initially have three (Sciaridae) or four (Cecydomyiidae) X chromosomes: two paternal X chromosomes and one or two maternal X chromosomes respectively. If both of the two paternally derived chromosomes are eliminated, embryos develop into males; if only one paternal or no X chromosome is eliminated, embryos develop into females. In scale insects (Coccoidea) that have an XX (female)/X0 (male) sex-determination system, all embryos develop from fertilized eggs (XX), but when the paternal chromosome becomes inactive through heterochromatinization, embryos (XO) develop into males (Nur, 1963, 1990). Both these systems entail the elimination of a paternal chromosome and that the maternal and paternal genomes act differently in sex determination, suggesting that the paternal and maternal genomes are epigenetically different. The molecular details of these paternal-specific chromosome elimination/inactivation systems are not known. My discovery of a parent-of-origin effect on a sex determination instructor gene suggests that the gene (s) involved in chromosome elimination in these other systems are maternally imprinted during oogenesis. Further studies are required to elucidate the nature of these imprinting mechanisms.

Another relevant case of a parent-of-origin-effect is the supernumerary (B) chromosome paternal sex ratio (PSR) in *N. vitripennis*. This paternally inherited chromosome causes complete paternal genome elimination after fertilization, except itself, resulting in male offspring of these haploidized eggs that contain PSR (Nur et al 1988; McAllister and Werren, 1997). A PSR located gene *haploidizer* has recently been identified in *N. vitripennis* to cause this effect (Benetta et al., 2020). *Haploidizer* encodes a putative protein consisting of a DNA-binding domain, that is predicted to target some core enzymes for histone posttranslational modifications resulting in the disruption of the paternal chromatin (Swim et al., 2012; Benetta et al., 2020) This is consistent with my results that the paternal genome is required for female

development, and the predicted function of *haploidizer* is also consistent with my finding that epigenetic modification is involved in sex determination in *Nasonia*.

This study is the first that provides clear molecular evidence for genomic imprinting involved in insect sex determination. Haplodiploidy, with diploid female and haploid male development, is a reproductive mode that occurs in about 15% of invertebrates, and the insect groups Hymenoptera and Thysanoptera are entirely haplodiploid (Kaiser and Bachtrog, 2010; Beukeboom and Perrin, 2014; Filia et al, 2015). A parent-of-origin effect in sex determination based on genomic imprinting may be a common theme in haplodiploids. Within the Hymenoptera, genomic imprinting may be involved in the sex determination of many non-CSD species. However, the sex-determination genes that are imprinted in those non-CSD species must be different from those of *Nasonia*, as *wom* is only present in *Nasonia* and *Trichomalopsis* species. At this point, it remains to be seen how widespread the MEGISD model applies to other hymenopterans and to what extent comparable mechanisms of genomic imprinting sex determination have evolved in haplodiploids.

Open questions about the molecular mechanism of imprinting and *tra* activation

What is the underlying mechanism for maternal imprinting of wom?

Our study reveals that *wom* is maternally imprinted, but the epigenetic mechanism underlying maternal imprinting of *wom* remains unknown. DNA methylation has been the most reported mechanism governing genomic imprinting in various eukaryotes over the past 20 years. It is the process that modifies the C-5 position of the cytosine of DNA by replacing it with a methyl group, typically resulting in repression of gene transcription (Bird, 1984). Thus, one cogent hypothesis is that the promoter region of *wom* is maternally methylated during oogenesis leading to the silencing of the maternal allele of *wom*. I observed that the *wom* promoter region is indeed enriched with CpG islands. However, in a pilot experiment, no DNA methylation was found in the *wom* promoter region of early male and female embryos. Furthermore, Wang et al. (2015) screened the genome-wide DNA methylation in adult males and females of *N. vitripennis* and *N. giraulti* and did not find significant DNA methylation differences between males and females. Regardless, it would still be worthwhile to examine DNA methylation in germline cells to confirm whether DNA methylation is responsible for maternal imprinting of *wom* in *Nasonia*.

An alternative hypothesis for *wom* silencing is histone modification, i.e. amino acids of the histone proteins are modified resulting in an alteration of the chromatin structure of the *wom* region during oogenesis preventing accessibility of the maternal *wom* allele. To test this possibility, chromatin immunoprecipitation sequencing (ChIP-seq) for profiling the histone

modifications with an antibody specific for the histone mark can be performed to compare the histone modification of the *wom* region between ovary and testis. In addition, as this histone modification is proposed to be exclusive to ovaries, the histone modifier gene should be a maternal effect gene that is specifically (significantly higher) expressed in ovary tissue. Yet, another promising approach towards unraveling the imprinting mechanism could start by comparing the transcriptome sequences between ovary and testis in *Nasonia*.

The gynandromorphy strain HiCD12 can also be very useful for investigating the mechanism underlying maternal imprinting of *wom*. No matter whether DNA methylation or histone modification is responsible for the maternal allele of *wom* repression, it should be different in HiCD12, and the sequence of the gene involved in DNA methylation or histone modification should differ from that of the wild-type. In addition, the cytoplasmic factor (e.g. mitochondria or small RNAs) that is involved in silencing of *wom* should also be different between in HiCD12 and wild-type. Thus, comparing the mitochondrial DNA and small RNA profile between HiCD12 and wild-type should be considered for future research.

How is wom involved in the timely activation of zygotic tra expression?

My study reveals that *wom* is involved in activation of zygotic *tra* expression, but does not provide adequate evidence that *wom*, directly or indirectly, regulates *tra*. *Wom* encodes a protein that contains a P53-like and a coiled-coil domain, suggesting that WOM may function as a transcription factor. If *wom* is the direct regulator of *tra*, WOM should bind to the promoter region of *tra* boosting *tra* transcription, consistent with the finding that a *tra* expression peak is present right after the *wom* expression peak. To test this possibility, the *tra* promoter region can be searched for any potential *wom* binding sites, and a ChIP assay can be conducted using a specific antibody for *wom* and followed by qPCR to examine if WOM indeed binds to *tra*. On the other hand, if *wom* does not bind to *tra*, *wom* must target another gene(s) that is required for activation of zygotic *tra*. Genome-wide ChIP-seq can be used to identify *wom* target gene (s). These possibilities would be worth investigating in future research to uncover how *wom* is involved in activating zygotic *tra*.

The origin of instructor genes in insects

All identified instructor genes in insects represent novelty in their gene sequences and evolutionary origins. Where they come from and how they evolve their function are important unanswered questions. In chapter 4, I presented evidence that *wom* is a novel instructor gene, which is not homologous to the only other currently known instructor gene of Hymenoptera, the *csd* gene (Beye et al., 2003), neither in sequence nor in function. *Wom* is only present in *Nasonia* and *Trichomalopsis* lineages. It contains a P53-like domain coding region and a partial duplication of LOC100678853, which is also only present in *Nasonia* and *Trichomalopsis* species and located close to *wom*. Phylogenetic analysis reveals that *wom* evolved from a *p53* homolog (*p53-2*) by incorporating a partial duplication of LOC100678853 before the split of *Nasonia-Trichomalopsis* and other species. This means that the instructor gene of *Nasonia* has recently evolved from existing genes, which is consistent with reports of other instructor genes, such as the *csd* in honeybee that evolved as a duplication of its downstream target gene *feminizer/tra* (Hasselmann et al., 2008) and *Mdmd* in housefly that originated from a duplication of its paralog, the splicing factor *CWC22* (Sharma et al., 2017). These are duplications of existing genes, but this does not seem to always be the case. The three M-factors, *MoY* (medfly *Ceratitis capitata*) (Meccariello et al., 2019), *Yob* (mosquito *Anopheles gambiae*) (Krzywinska et al., 2016), and *Guy1* (mosquito *Anopheles stephensi*) (Criscione et al., 2016), show no homology to any existing genes. They are all short and unique sequences, suggesting that these three instructor genes have originated *de novo* from ancestral non-coding sequences (pre-existing). My work, in combination with the currently available studies, indicates that instructor genes in insect sex determination evolve rapidly by gene duplication and genomic rearrangements. Future investigation of molecular mechanisms that lead to gene duplication and genomic rearrangements may help to understand the evolutionary forces driving the novelty of instructor genes in insects. Overall, the diversity and rapid evolution of instructor genes remain an intriguing field of study.